SANDIA REPORT
SAND2003-3158
Unlimited Release
Printed September 2003

Dual Mode Use Requirements Analysis for the Institutional Cluster

Robert W. Leland

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2003-3158
Unlimited Release
Printed September 2003

Dual Mode Use Requirements Analysis for the Institutional Cluster

Robert W. Leland
Computer and Software Systems
Sandia National Laboratories
P. O. Box 5800, MS-0310
Albuquerque, NM 87185-0310

Abstract

This paper analyzes what additional costs would be incurred in supporting dual-mode, i.e. both classified and unclassified use of the Institutional Computing (IC) hardware. The following five options are considered: *periods processing* in which a fraction of the system alternates in time between classified and unclassified modes, *static split* in which the system is constructed as a set of smaller clusters which remain in one mode or the other, *re-configurable split* in which the system is constructed in a split fashion but a mechanism is provided to reconfigure it very infrequently, *red/black switching* in which a mechanism is provided to switch sections of the system between modes frequently, and *complementary operation* in which parts of the system are operated entirely in one mode at one geographical site and entirely in the other mode at the other geographical site and other systems are repartitioned to balance work load. These options are evaluated against eleven criteria such as disk storage costs, distance computing costs, reductions in capability and capacity as a result of various factors etc. The evaluation is both qualitative and quantitative, and is captured in various summary tables.

Acknowledgements

Many people helped in the preparation of this whitepaper. I would like to thank Jim Tomkins, Bill Camp, Doug Doerfler, Pete Dean, Frank Mason, John Noe, Paul Iwanchuck and Jim Ang for reviewing drafts and making valuable suggestions. I would also like to thank:

- Doug Doerfler, Ron Brightwell and Jim Laros for providing substantial content for sections 3A through 3E, the issues analysis on disk storage costs, management hardware costs, interconnect hardware costs, diskless operation and security approval.

- Mike McGlaun and John Zepper for their thinking on disk storage capacity needs.

- Phil Heerman for content on section 3F on distance computing costs.

- John Noe for advice and draft text on section 3K on operational complexity.

- Jim Ang for analysis on the breakdown of classified and unclassified computing and general support across many phases of the mid range computing effort.

- Steve Kleban and Scott Clearwater for running the queuing simulation studies summarized in section 3I and 3J, and to Kathy Hiebert-Dodd for contributing her time to the effort.

- Vitus Lueng and Jeanette Johnston for running a parallel simulation study and theoretical analysis, and to David Womble for reviewing their study and contributing their time.

- Bill Camp, Jim Tomkins, Mike McGlaun and John Noe for their thinking downtime valuation and restart frequency.

Contents

(This page left intentionally blank)

SAND2003-3158
Unlimited Release
Printed September 2003

Dual Mode Use Requirements Analysis for the Institutional Cluster

Robert W. Leland
Computer and Software Systems
Sandia National Laboratories
P.O. Box 5800, MS-0310
Albuquerque, NM 87185-0310

0. Executive summary

This paper analyzes what additional costs would be incurred in supporting dual-mode, i.e. both classified and unclassified, use of the Institutional Computing (IC) hardware. Several assumptions are made:

- The volume of classified and unclassified work is approximately equal, an assumption well supported by the available data.
- The cost of operating a given machine entirely in one mode is the same as that of operating it entirely in the other.
- In each case the CA machine operates in one mode permanently, the consensus being that the size of this system is not sufficient to amortize the effort and cost of dual-mode use.

Five options are considered.

a) **Periods processing:** The NM machine alternates in time between classified and unclassified modes.

b) **Static split:** The NM machine is constructed as two smaller clusters. Here it is assumed these are of equal size (512 processors), and that this division would be maintained permanently.

The work described in this report was performed for Sandia National Laboratories under Contract No. DE-AC04-94AL85000.

c) **Re-configurable split:** The NM machine is constructed in a manner that it can be reconfigured occasionally e.g. up to a few times per year to achieve different ratios of classified to unclassified computing.

d) **Red/black switching:** The NM machine has a center section connected alternately in time to classified and unclassified end sections ("heads") in a manner that allows relatively frequent switching, e.g. every few weeks.

e) **Complementary operation:** The NM machine is operated entirely in one mode and the CA machine is operated entirely in the other mode.

These options are evaluated against the eleven issues listed in Table 0 below. For each issue/option pair in which there was judged to be a potential impact, an estimate of the range of this impact was developed, and a point within this range was selected.

Quantitative impact estimate by issue	Opt. A Periods	Opt. B Static	Opt. C Reconfig.	Opt. D Red/Black	Opt. E Comp.
Disk storage costs increased	**0k**	**0k**	**0k**	**0k**	**0k**
Management HW costs	**35k**	**5K**	**35k**	**35k**	
Interconnect HW costs increased			**0k**	**70k**	
Diskless operation required	**Reduced to 2 Vendors**		**Reduced to 2 Vendors**	**Reduced to 2 Vendors**	
Security approval delay	**1 month**		**1 month**	**1 month**	
Distance computing required					**0k**
Cplant costs induced					**0k**
Capability reduced		**To ½ of maximum**	**To ½ max in normal ops**	**To ¾ max in normal ops**	
Capacity reduced		**4% = $400K**	**1% = 100K**		
Response time (average across jobs)		**~2X worse: from 7.5hrs to 16.7hrs**			
Operational complexity increased	**Increased**		**Increased**	**Increased**	

Table 0. Quantitative impact estimate by issue (Table 2 in text below.)

Options A (periods processing), C (re-configurable split), and D(red-black switching) all have minor capital or operational cost impact. For example, the largest cost, that for Red-Black switching, was estimated at just over $100k, about 1% of the estimated lifecycle cost of the New

Mexico system. However each of these options would likely induce a moderate delay (order a month) in schedule, and would require the procurement of a diskless cluster. It is likely that this would restrict the field of willing bidders, perhaps to just 2 or 3 companies with experience fielding diskless systems. Hence these options present the greatest risk to competitive bidding and timely delivery of a system.

Option B (static split) reduces the capability of the New Mexico system to ½ its potential capability. It is projected to reduce the capacity of the New Mexico machine by approximately 4% (5% loss in scheduling efficiency, 1% improvement due to increased reliability) and this is valued at about $400k over the lifecycle cost of the machine (estimated at $10M). It is projected to increase the average wait time for jobs by a factor of 2.2 from 7.5 hours to 16.7 hours. Hence this option has the largest impact on machine performance.

Option E (complementary operation) is projected to have no impact on system delivery, cost or performance, but raises policy issues. Under this option, the volume of cycles provided to classified and unclassified users would be the same as in the other options, but the source for some of these cycles would likely include other systems, e.g. Cplant. Hence this would raise the issue of whether consumers of institutional computing cycles cared which platform provided these cycles. It would also likely imply a commitment to keep Cplant operational until compensating resources were provided e.g. an expansion of the Institutional Cluster network, or some portion of Red Storm. Hence this option is the most complex from a policy perspective.

1. Introduction

The purpose of this paper is to analyze what additional costs would be incurred in supporting dual-mode use of the IC hardware and what additional technical specifications would be necessary to support various options for implementing dual-mode use.

Several assumptions are made:

- The volume of classified and unclassified work is approximately equal, an assumption well supported by the available data.

- The cost of operating a given machine entirely in one mode is the same as that of operating it entirely in the other.

- In each case the CA machine operates in one mode permanently, the consensus being that the size of this system is not sufficient to amortize the effort and cost of dual-mode use.

Five options are considered.

f) **Periods processing:** The NM machine alternates in time between classified and unclassified modes.

g) **Static split:** The NM machine is constructed as two smaller clusters. Here it is assumed these are of equal size (512 processors), however it is believed the induced cost is not very sensitive to this ratio. This division would be maintained permanently and no reconfiguration would be contemplated.

h) **Re-configurable split:** The NM machine is constructed in a manner that it can be reconfigured occasionally e.g. up to a few times per year to achieve different ratios of classified to unclassified computing. In some cases switching capabilities built into the COTS interconnect hardware may be sufficient to support the required duty cycle. If not, an inexpensive switching mechanism, e.g. paddle connectors may be sufficient. In either case it is expected that re-configuring the machine under this option would be more labor intensive and error prone than if traditional Red/black switching ala ASCI Red or Red Storm were used.

i) **Red/black switching:** The NM machine has a center section connected alternately in time to classified and unclassified end sections ("heads") in a manner that allows relatively frequent switching, e.g. every few weeks. This implies a more robust switching mechanism than in the option C, e.g. one like that in ASCI Red or Red Storm, would be required for reliable operation.

j) **Complementary operation:** The NM machine is operated entirely in one mode and the CA machine is operated entirely in the other mode. Given the assumption of equal classified and unclassified work loads, this option would likely require reconfiguration of other resources, for example Cplant, to compensate for the discrepancy in size between the NM and CA systems.

Eleven issues are evaluated:

A. Disk Storage costs (page 6)
B. Management hardware costs (page 7)
C. Interconnect hardware costs (page 7)
D. Diskless operation (page 9)
E. Security approval (page 10)
F. Distance computing (page 10)
G. Cplant costs (page 10)
H. Capability available (page 10)
I. Capacity available (page 11)
J. System response time (page 11)
K. Operational complexity (page 13)

These are discussed in detail in the lettered subsections of section 3, entitled *Issues Analysis*.

2. Summary of Impact

The following chart summarizes the analysis by issue in section 3 below qualitatively.

Qualitative impact	Opt. A Periods	Opt. B Static	Opt. C Reconfig.	Opt. D Red/Black	Opt. E Comp.
Disk storage costs	Neutral	Neutral	Neutral	Neutral	Neutral
Management HW costs	Increased	Increased	Increased	Increased	
Interconnect HW costs		Neutral	Increased	Increased	
Diskless operation	Required		Required	Required	
Security approval	Delayed		Delayed	Delayed	
Distance computing					Required
Cplant costs					Neutral
Capability available		Reduced	Reduced		
Capacity available	Negligible	Reduced	Reduced	Negligible	
Response time		Degraded			
Operational complexity	Increased		Increased	Increased	

Table 1. Qualitative impact assessment

Quantitative impact estimate by issue	Opt. A Periods	Opt. B Static	Opt. C Reconfig.	Opt. D Red/Black	Opt. E Comp.
Disk storage costs increased	0k	0k	0k	0k	0k
Management HW costs	35k	5K	35k	35k	
Interconnect HW costs increased			0k	70k	
Diskless operation required	Reduced to 2 Vendors		Reduced to 2 Vendors	Reduced to 2 Vendors	
Security approval delay	1 month		1 month	1 month	
Distance computing required					0k
Cplant costs induced					0k
Capability reduced		To ½ of maximum	To ½ max in normal ops	To ¾ max in normal ops	
Capacity reduced		4% = $400K	1% = 100K		
Response time (average across jobs)		~2X worse: from 7.5hrs to 16.7hrs			
Operational complexity increased	Increased		Increased	Increased	

Table 2. Quantitative impact estimate by issue

Quantitative impact range by issue	Opt. A Periods	Opt. B Static	Opt. C Reconfig.	Opt. D Red/Black	Opt. E Comp.
Disk storage costs increased	**0k-625k**	**0k-625k**	**0k-625k**	**0k-625k**	**0k-625k**
Management HW costs	**28k-50k**	**3k-10k**	**28k-50k**	**28k-50k**	
Interconnect HW costs			**0k-10k**	**30k-300k**	
Diskless operation required	**1 – 3 vendors**		**1 – 3 vendors**	**1 – 3 vendors**	
Security approval delay	**2wks. – 2mos.**		**2wks. – 2mos.**	**2wks. – 2mos.**	
Distance computing required					**0k-250k**
Cplant costs induced					**0k – neg. 100k**
Capability reduced		**To ½ of maximum**	**To ½ max in normal ops**	**To ¾ max in normal ops**	
Capacity reduced		**4% red. = $300-$500k**	**1% red. = $75k-125k**		
Response time		**2.2X worse: from 7.5hrs to 16.7 hrs.**			
Operational complexity increased	**Increased**		**Increased**	**Increased**	

Table 3. Quantitative impact range by issue

3. Issues analysis

A. Disk storage costs

Two categories of external storage are specified for the IC – 5Terabytes of disk to support the serial file system, and 20 Terabytes of disk to support the parallel file system. The following comments apply to all options listed above.

<u>5TB Disk Storage Serial File System</u>
The 5 Terabytes of disk storage would be used for the user home file system, to store, for example, code executables and input files. Its size was specified based on experience with similar systems and the expected number of users on the IC. Were users to maintain both

classified and unclassified accounts, their desire for storage would likely increase. However experience has shown that users will tend, collectively and over time, to fill all available disk space. Hence an effective purging policy is essential to rationally assessing the appropriate investment in disk space. Above some reasonable threshold, it is more a matter of convenience than necessity.

If we hypothesize that both classified and unclassified work-sets must or should be performed, the lab must or should purchase (or already has purchased) a proportionate amount of disk somewhere in the overall computing system to support this. This reasoning leads to one bound in the assessment – that no additional disk of this type is required for dual mode use of the IC. To arrive at the other bound we might consider the case in which the classified and unclassified users are orthogonal groups with completely independent needs and hence the entire disk system must be replicated.

A reasonable intermediate case would assume there is significant overlap between these groups and their needs, and that to support a second set of accounts users might on average want 50% more storage on the system. This would suggest an extra 2.5 Terabytes of disk on the serial file system.

The most likely scenario is that we would initially purchase no additional disk, leaning strongly towards the first perspective and arguing that, since the disk system will not in any event fill up immediately, and that disk prices have dropped consistently over time, we can afford to defer the decision, and will actually benefit from doing so.

<u>20TB Parallel File System Disk</u>

The 20 Terabytes of disk storage would be used for the high performance parallel file system, to store typically restart files and large data sets in transit to visualization and data analysis servers. Hence this file system will be treated as a scratch space and not backed up. Its size was specified based on experience with similar systems and the expected computational rate and memory capacity of the IC, representing about the same storage capacity relative to computational capacity as that of Red Storm. Since the NM machine will have approximately a Terabyte of memory, it would not be difficult to fill the entire parallel file system, and here again an effective management policy that motivates users to behave in a disciplined fashion is essential.

At one extreme we would argue that to support dual-mode use, we would need to completely replicate the disk system, adding 20 Terabytes of disk. At the other we would argue that if use of the parallel file system is dominated by restart files we can afford to split the disk into a red 10 Terabyte subsystem and a black 10 Terabyte subsystem because each of these can support writing the full state of the memory 10 times over, and that should be sufficient to store the last copy of a restart file capturing the state of an application. In this case we need no extra disk.

An intermediate argument would be that if we decide upon traditional Red/Black switching and break the machine into, say, a ¼ red head, a ¼ black head, and a ½ center section, we would need an additional 10 Terabytes (or 50%) to replicate the disk for the center section added memory.

Periods processing constitutes another potentially intermediate case. If 10 Terabytes is sufficient in one color, then it should also be sufficient in the other and hence if we had run the machine entirely in one mode we would in fact have needed only 10 Terabytes. That is to say, perhaps we do in fact need to double the disk to accommodate dual mode use, but the original spec of 20 Terabytes is too generous and masks this. If we really only need 2.5 Terabtyes and 10 Terabytes respectively, by purchasing 25 Terabytes, we are therefore incurring the additional cost of 12.5 Terabytes.

Collating these arguments we find that the range of possible additional disk required is 0 to 25 Terabytes, a reasonable intermediate position would be that 12.5 Terabytes of this is tacitly covering the demands of dual mode use, and our likely conclusion is that we need no additional disk (at least immediately) to cover dual mode use.

Noting the cost range for disk is $15k to $50k per Terabyte, and estimating that we would arrive at a price point of about 25K per Terabyte, this suggests an induced cost in the range of $0k to $625k, and that we will list this as an induced cost of $0k by the arguments given.

B. Management hardware costs

Each of the clusters will have a management workstation that serves as a single point of control for the system. Since this workstation will have associated disk storage, the same hardware cannot be used in both classified and unclassified modes as that would allow transferring of classified data into the unclassified environment. Hence options A (periods processing), C (re-configurable split) and D (red black switching) would require the addition of a second, physically distinct management workstation to the New Mexico cluster. Option B (static split) would also require this second workstation in order to support administration of a second cluster. The management workstation could in principal range between a low-end pc running a variant of windows and costing in the neighborhood of $3k, to a high end Unix workstation costing in the neighborhood of $10k. The midpoint of this range, $5k, is chosen here as the cost estimate.

A second management workstation will in practice require an independent network capability in the case of options A (periods processing), C (re-configurable split) and D (red black switching). It is assumed here this will be Ethernet at a cost of between $25k and $40k to connect to 1024 processors, assuming one inexpensive processor would be used for every 32 processors in the system. The estimate used here for the cost is $30k.

Combining theses estimates, the expected range for options A (periods processing), C (re-configurable split) and D (red black switching) is between $28k and $50k, with an estimated cost of $35k. The expected range for option B (static split) is $3k to $10k with an estimated value of $5k.

C. Interconnect hardware costs

Interconnect hardware is meant here to apply to the primary, high speed interconnect designed to carry message traffic consisting of application data. In options C (re-configurable split) and D (red black switching) some mechanism must exist for modifying the network topology of the machine. In the case of option C (re-configurable split) it is expected that this would occur very infrequently, say at most a few times per year, and hence the mechanism need only be robust and convenient to that duty cycle. Hence a paddle connector system like that used in Cplant would probably be sufficient and it is estimated roughly that the induced hardware cost here would be low, less than $10k. The current generation of CLOS network switches available from a leading vendor are designed to build large networks in a modular fashion and hence have a limited switching capability built in. Optical fiber lines used to connect switching units are terminated with connectors that can be plugged into other interconnect fabric units and subsequently unplugged without damage at a frequency that is probably sufficient for this option. However these connectors are not designed for this, and using it in this operational mode, particularly if the switching frequency were to creep up, is considered a significant reliability risk likely to induce indirect costs (discussed below). Hence the induced hardware cost in this case may be as low as $0k. For option C (re-configurable split) we estimate the range at $0k - $10k with a likely direct cost of $0k.

In the case of option D (red black switching), a more robust and convenient switching mechanism is required to support switching on a frequency potentially measured in weeks or even days. Vendors are reluctant to discuss specific costs for this capability outside the context of a larger bid, but we can note that all of the very high quality red-black switching hardware to be incorporated in Red Storm is expected to cost less than $40k, hence this seems a reasonable upper bound for the required hardware for the smaller Institutional Cluster. However engineering would probably be required to insert these connectors into the cables and to perform signal integrity and other validation tests. Estimating this effort at three months of an engineer's time we arrive at an additional cost of approximately $60k. Hence the upper bound would be approximately $100k. A lower bound of $10k for hardware was arrived at somewhat arbitrarily since this matches the upper bound of the estimated cost for the cheaper paddle connector approach. Assuming the additional engineering could be completed in one month, we add $20k to arrive at a lower bound of $30k. A mid point of $70k was selected as the estimate.

One vendor gave a rough quote of 20-40% of switch fabric costs for this incorporating Red/Black switching. Assuming they were to bid a single rail 512 dual-processor node system and using the switch fabric cost quoted us as a direct purchaser ($1M), these costs could go as high as $400k. However this quote was considered very conservative based on the particular vendor's inexperience with this issue. To arrive at the upper end of the range estimate the 30% figure was applied to arrive at $300k, and the reasoning in the previous paragraph was used in arriving at the point estimate of the cost.

In the case of option C (re-configurable split) there is also the possibility that two 512 processors systems would require less interconnect fabric than one 1024 processor system which would mean there would actually be a cost reduction associated with interconnect hardware. In fact, in the case of Myricom, the price of these switches is exactly proportional between the two cases because the additional spine layer of switching is required in both, hence there is no pricing advantage to splitting the system arising in this context. Quadrics conceded that there might be some price advantage to splitting the system, but would not quantify this. Since Myricom appears the most likely source for interconnect technology based on it's current market share, we assumed here no savings in switch fabric costs would arise from option C (re-configurable split).

Another significant cost associated with these switching options may well be embedded in their impact on reliability. Here it is assumed that in case of option C (re-configurable split) we will use connectors built into the switching fabric and which are hence reasonably reliable. However the operational difficulty of verifying correct wiring and proper electrical connectivity is substantially higher with poorer quality connectors. Switching under option C (re-configurable split) might require order a day, whereas under option D (red black switching) changing and verifying the configuration might typically require order and hour. Assuming three changes per year, a lost day with each, and 300+ active days per year of computing, this is order a 1% affect on capacity.

In the case of option D (red black switching), we assume high quality connectors with adequate reliability are used, and the reliability related switching costs are therefore negligible.

D. Diskless operation

Typically large clusters are sold with disk drives included in each compute node. This is problematic in the case of options A (periods processing), C (re-configurable split) and D (red black switching) where a compute node must operate in both classified and unclassified modes since, absent an elaborate disk scrubbing procedure, this would permit the transfer of information between environments. Removing these disks should in principal reduce cost and improve reliability provided a vendor has software capable of booting the machine without disks. However, very few vendors have experience standing up diskless machines, hence the primary impact of this requirement would be to reduce the number of viable vendors. One of the expected vendors has delivered a diskless system (Cplant) to us before, and a second has delivered another diskless system to another site. Hence we estimate the number of vendors who could credibly bid a diskless system is between 1 and 3, and we estimate that two would do so. This would substantially reduce the level of competition in the process, particularly if one of these two vendors had difficulty meeting another requirement and that became known to its competitor.

E. Security approval

It is expected that in the case of options A (periods processing), C (re-configurable split), and D (red black switching), the additional complexity of security operations will induce a delay of between a few weeks and several months into the over-all schedule. An estimate of one month is used below to quantify the likely impact. This delay is associated primarily with classified processing although some dedicated time on the system would be needed to perform security tests.

F. Distance computing

Option E (complementary operation) would require adequate support for distance computing across the sites. The consensus is that support for classified computing at a distance over SecureNet would be effective, and distance computing support for unclassified usage would be more limited in data rate but probably adequate with the existing infrastructure. A higher estimate for the cost to upgrade infrastructure is $250k, and this is taken as the upper bound in the tables below. A hedge against this downside risk would be to provide no additional infrastructure until some operational experience were gained since this would not preclude reconfiguring the machine into two separate clusters were it decided that the infrastructure costs were unacceptable.

G. Cplant costs

The likely scenario driving this issue is that we would operate the IC entirely in unclassified mode and repartition more of Cplant into classified mode to compensate. Our experience has been that high performance computing systems (including Cplant) operate more reliably with fewer users and codes and more dedicated time, all characteristics of classified usage. Hence it is expected that Cplant operation costs could be reduced under option E (complementary operation), perhaps by as much as half an FTE valued at $200k for a typical support staff member. Because the mission importance and urgency of classified work may on average be greater than that of unclassified work, it seems more likely that we would choose instead to maintain the same level of support for the machines and provide relatively greater and more timely support to the typical user running in classified mode. Hence the estimate used here is that the cost might decline as much as $100k, but likely there would be little to no impact on Cplant costs.

H. Capability available

Capability would be reduced to half of the full 1024-processor capability in case of option B (static split, assumed to b 2x512 processors). In practice option C (re-configurable split) is the same. Option D (red black switching) would limit capability to ¾ of the maximum in the ¼, ½, ¼ configuration used in Red, but this could be avoided with an additional set of switches as deployed in Red Storm, and these additional layers of switching was assumed in the costing of switching hardware discussed earlier. However, in normal operations we would expect the machine to be configured as Red is, and therefore capability would be reduced.

It is not clear how to value loss of capability. One approach to valuing the tangible costs associated with lost capability is to assume that the larger jobs which might have run on the 1024 processor system (namely 512 processor jobs under the assumption queing policy would normally limit jobs to half the machine) will now be run on Red Storm. Since Red Storm is a more expensive machine per peak flop rate, there may be some additional cost associated with this work (or there may not be if the greater parallel efficiency of Red Storm outweighs this effect) for the job size run. In practice it seems likely the main effect would be to shift the work profile on Red Storm, drawing it more into the capacity range than would otherwise be the case and adding to it’s congestion. This may also have intangible costs to the ASCI program since the rational for Red Storm is based substantially on capability need.

I. Capacity available and J. Response time

The Kleban/Clearwater (Org. 6544) study commissioned for the IC is summarized in Table 1 included below. This analysis used the processor count and job length distribution observed on Cplant (up to 1024 processor jobs) run through a queuing simulation study and estimated a 5.1% drop in throughput upon dividing a 1024 processor system into two 512 processor systems. Jobs requiring more processors than available were taken out of this throughput calculation.

Average wait time for jobs more than doubled as a result of partitioning, from 7.5 hours to 16.7 hours. The "expansion factor" (EF) – the average turnaround times for jobs normalized to their run time – increased by a factor of more than 2.5 from 929 to 2348. This EF is heavily influenced by small jobs that are highly delayed in a relative sense, but not an absolute sense, so another useful figure of merit is the EF for sizeable jobs – say those requiring more than an hour to execute. This increased from 2.9 to 3.5, a roughly 20% increase. It is likely that these full affects would not be observed in actuality because users would adapt their behavior to compensate, but they do suggest a significant increase in congestion and decrease in system responsiveness from the user perspective upon partitioning.

A second simulation study performed by Leung (Org. 9215) reached a similar conclusion and correlated this with mathematical queuing theory.

Table 1—Performance Parameters

Config-uration n x Part.	EF (avg. / median)	EF (avg./ median) > 1 hour	Wait (sec) (avg. / median)	Usage	Frac Throughput
1 x 1024	929 / 6	2.9 / 1.3	27k / 916	.834	1.000
2 x 512	2348 / 52	3.5 / 1.6	60k / 6959	.782	.949
4 x 256	3303 / 38	4.7 / 1.8	98k / 7786	.758	.921
8 x 128	3143 / 21	5.3 / 1.8	100k / 2546	.696	.928

where

$$\mathrm{EF} = \frac{finish\ time - submit\ time}{runtime} = \frac{waittime + runtime}{run\ time} = 1 + \frac{wait}{run}.$$

A second potentially significant effect on capacity relates to the impact on the machine reboot rate of splitting the machine into smaller configurations. To assess this, we consider two classes of machine failures:

- Failures that are local in impact and hence their impact is proportional to the number of parts and no significant difference arises between the two cases.

- Failures that have a global impact and require at least a full reboot to fix. Here there could well be a difference since a single error takes out more capacity in a larger system.

Say the global errors occur with frequency x in a *512* processor system. So the chance that system is up over some given interval of time is *(1-x)*. If a second *512* system is independent, the chance that is up is also $(1-x)$. The chance they are both up is hence $(1-x)^2$. The chance that it is down is therefore $1-(1-x)^2 = 2x - x^2$. Assuming x is small, this is approximately $2x$. So in the *2* by *512* case we lose over some period of time $512*x + 512*x$ processors, and in the *1024* case we lose $1024*2x$. Hence the lost capacity in the *1024* case is twice the lost capacity in the *2* by *512* case. This is intuitive, because under the assumptions, whenever we would have lost one machine in the *2* by *512* case, we also lose the other because they are always tied together. The naturally arising questions are

- What is the global failure rate, x?
- What is the impact on capacity of these global failures in the two cases?

A reasonable estimate for the global failure rate on a 1024 processor cluster is once every 100 hours since this is approximately the reboot for a similar size configuration of Cplant.

Since the machine can be rebooted quickly, the main impact on capacity is the lost work rather than the down time. Lost work is a function of the rate at which restart files are written by the application codes. No historical record exists that would provide data on this, and discussions with various applications teams indicate that restart frequency varies quite widely with code, specific application of the code, and personal practice. For example, CTH runs often dump restart files every 1 to 2 hours, whereas Salinas is often run for order 24 hours without restarting. Furthermore, the restart frequency can be varied within certain constraints of efficiency to match a machine environment. The estimate used here is that the average restart file frequency is every 4 to 6 hours, Mike McGlaun's estimate for the ASCI applications. Since many jobs on the system complete in less than this time and hence have in effect shorter restart frequencies, the lower end of this range (every 4 hours) was used. Assuming reboots are randomly distributed in time, the expected lost work then is 2 hours (since some restarts occur 4 hours from the last restart, but an equal number occur 0 hours from it). Since this occurs every 100 hours by hypothesis, the lost capacity is approximately 2% in the 1024 processor case. By the algebraic argument given above, the lost capacity in the 2 by 512 case is one half this, or 1%. Hence the differential is 1% of system capacity in favor of the 2 by 512 configuration.

Summarizing, by splitting a 1024 processor machine into 2 machines with 512 processors, we expect to lose about 5% capacity in queing efficiency and to gain about 1% capacity in reliability for a net effect of 4% loss in capacity. To value this we note that lifecycle costs for linux clusters were estimated in the business case for this class machine to be approximately 2.4 times the hardware costs. Assuming some economies arising from operating both the 256 processor machine in CA and the 1024 processor machine in NM, the life cycle cost for both machines is predicted to be approximately \$10M. Hence a 4% loss of capacity over a 6 year lifecycle of the machine would be valued at approximately \$400k. Another estimate on this is arrived at using the \$1.6M per year operational costs currently proposed as the amount to be recovered per year over a 5 year period in the Institutional Cluster rate structure. This would put the lifecycle costs for the full suite of machines at \$12.6M (\$4.6M capital and \$8.0M operating). Making the simplifying assumption that the operational costs are proportional to the node count, 80% of this \$12.6M, or \$10M would be the lifecycle cost of the New Mexico machines. Hence here again the 4% the lost capacity would therefore be valued at \$400k. We do not have error bars for the results of the simulation study or the operational costs, so to arrive at the range estimate a rather arbitrary figure of 25% error in these estimates was used. Hence the range estimate for this 4% loss of capacity is $300k – $500k.

Kleban and Clearwater performed a second simulation study in which rather than truncating jobs using more than 1024 processors from the Cplant workload distribution, they mapped this into jobs that would run on fewer processors for a longer period. This alters the processor and runtime distributions and results in a relatively flat distribution of throughput upon subdivision of the machine until the 8 by 128 configuration, assuming a common queuing system is provided. However, to fairly compare these scenarios the remapped job distribution should be re-run on the larger configuration, achieving improved throughout due to packing efficiency. Hence in the simulation performed the larger configurations are being penalized with respect to utilization by supporting a more capability rich workload. The assumption used here in arriving at the point estimate is that jobs requiring more processors than available would migrate to a capability platform. That allowed use of the study better matched to the situation in question and did not require the assumption that the remapped workload was feasible or representative.

J. Response time

This subject is treated in the previous section.

K. Operational complexity

Static configurations of systems present the least operational complexity. The benefit derived from identical system configurations is an improved administration process and an ability to prototype new operating system or runtime features on one (hopefully unclassified) system prior to committing it to all systems. System upgrades are voluntary in many cases but required for some needed application improvement or security patch and can be scheduled to limit customer impact in any configuration. Periodic operating system upgrades are a natural occurrence and not of major impact to the configuration options.

Switchable configurations, in whatever variant, introduce configuration complexities, additional operational procedures to preclude data transfer between security modes, a two person involvement during reconfiguration, and utilization reporting complexity. In a well developed system environment, these impacts are minimal and present no additional burden to administration but do present scheduling issues that affect customer expectations. Scheduling changeovers and accommodating customer needs has a larger impact than the physical reconfigurations. There is some small possibility that switching will introduce unexpected instabilities due to variations in system load or some lack in system administration procedure.

Periods processing introduces reconfiguration and security complications that require additional support. Estimates of the cost impact of this additional effort are dependent on the specific methodologies employed to convert from one security environment to another. Manual reconfiguration of network attachments and power cycling if necessary can have a significant cost in lost compute node availability and decreased hardware reliability.

Complementary operation with other resources to balance workloads is achievable with little impact to support needs. Somewhat more involved scheduling and coordination activities would be required but are well within the normal work requirements for large system administrators.

4. Summary evaluation

Options A (periods processing), C (re-configurable split), and D(red-black switching) all have minor capital or operational cost impact. For example, the largest cost, that for Red-Black switching, was estimated at just over $100k, about 1% of the estimated lifecycle cost of the New Mexico system. However each of these options would likely induce a moderate delay (order a month) in schedule, and would require the procurement of a diskless cluster. It is likely that this would restrict the field of willing bidders, perhaps to just 2 or 3 companies with experience fielding diskless systems. Hence these options present the greatest risk to competitive bidding and timely delivery of a system.

Option B (static split) reduces the capability of the New Mexico system to ½ its potential capability. It is projected to reduce the capacity of the New Mexico machine by approximately 4% (5% loss in scheduling efficiency, 1% improvement due to increased reliability) and this is valued at about $400k over the lifecycle cost of the machine (estimated at $10M). It is projected to increase the average wait time for jobs by a factor of 2.2 from 7.5 hours to 16.7 hours. Hence this option has the largest impact on machine performance.

Option E (complementary operation) is projected to have no impact on system delivery, cost or performance, but raises a policy issue. Under this option, the volume of cycles provided to classified and unclassified users would be the same as in the other options, but the source for some of these cycles would likely include other systems, e.g. Cplant. Hence this would raise the issue of whether consumers of institutional computing cycles cared which platform provided these cycles. It would also likely imply a commitment to keep Cplant operational until compensating resources were provided e.g. an expansion of the Institutional Cluster network, or some portion of Red Storm. Hence this option is the most complex from a policy perspective.

DISTRIBUTION:

1	MS-0139	Mike Vahle, 9900
1	MS-0139	Robert K. Thomas, 9904
1	MS-0321	W. J. Camp, 9200
1	MS-0310	Jim Ang, 9220
1	MS-0310	R. W. Leland, 9220
1	MS-0310	M. D. Rintoul, 9212
1	MS-0316	John Aidun, 9235
1	MS-0316	Sudip Dosanjh, 9233
1	MS-0318	Jennifer Nelson, 9209
1	MS-0318	Paul Yarrington, 9230
1	MS-0801	Frank Mason, 9320
1	MS-0807	John Noe, 9328
1	MS-0819	Randy Summers, 9231
1	MS-0820	Pat Chavez, 9232
1	MS-0822	Phil Heerman, 9227
1	MS-0823	John Zepper, 9324
1	MS-0835	Mike McGlaun, 9140
1	MS-0847	Ted Blacker, 9226
1	MS-0847	Scott Mitchell, 9211
1	MS-1109	Paul Iwanchuk, 9220
1	MS-1109	Jim Laros, 9224
1	MS-1109	James L. Tomkins, 9220
1	MS-1110	Ron Brightwell, 9224
1	MS-1110	Doug Doerfler, 9224
1	MS-1110	Jeanette Johnston, 9223
1	MS-1110	Vitus Leung, 9215
1	MS-1110	Neil Pundit, 9223
1	MS-1110	David Womble, 9214
1	MS-1111	Bruce Hendrickson, 9215
1	MS-1137	Kathie Hiebert-Dodd, 6545
1	MS-1137	Steve Kleban, 6544
1	MS-9018	Central Technical Files, 8945-1
2	MS-0899	Technical Library, 9616